

# **JP2001175807**

Publication Title:

METHOD FOR SELECTING TEXT AREA

Abstract:

Abstract of JP2001175807

**PROBLEM TO BE SOLVED:** To provide a method for selecting a text area from an image. **SOLUTION:** Characters by words in an image and format information are accessed from a memory device, and active areas made to correspond to the respective words in the image based upon the characters and format information are determined; and the initial and final words in a text area selected on the basis of at least one active area made to correspond to at least one word in the selected area are discriminated and on the basis of the initial and final words in the selected text area, all the words in the selected text area are discriminated.

Data supplied from the esp@cenet database - Worldwide

---

Courtesy of <http://v3.espacenet.com>

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2001-175807  
(P2001-175807A)

(43)公開日 平成13年6月29日 (2001.6.29)

(51) Int.Cl. <sup>7</sup> G 0 6 K	9/20	識別記号 3 4 0	F I G 0 6 K	9/20	テ-マコード <sup>*</sup> (参考) 3 4 0 L 3 4 0 B
	9/32			9/32	
G 0 6 T	7/40	1 0 0	G 0 6 T	7/40	1 0 0 A

審査請求 未請求 請求項の数 2 O L (全 14 頁)

(21)出願番号	特願2000-335284(P2000-335284)
(22)出願日	平成12年11月2日 (2000.11.2)
(31)優先権主張番号	4 3 8 8 9 1
(32)優先日	平成11年11月12日 (1999.11.12)
(33)優先権主張国	米国 (US)

(71)出願人	590000798 ゼロックス コーポレーション XEROX CORPORATION アメリカ合衆国 コネティカット州・スタ ンフォード・ロング リッチ ロード・ 800
(72)発明者	マウリティウス ゼーガー イギリス国 エスジー8 7エスユー ニ アー ロイストン フаールミア ハイ ストリート フаールミア フаールミア ハウス
(74)代理人	100079049 弁理士 中島 淳 (外1名)

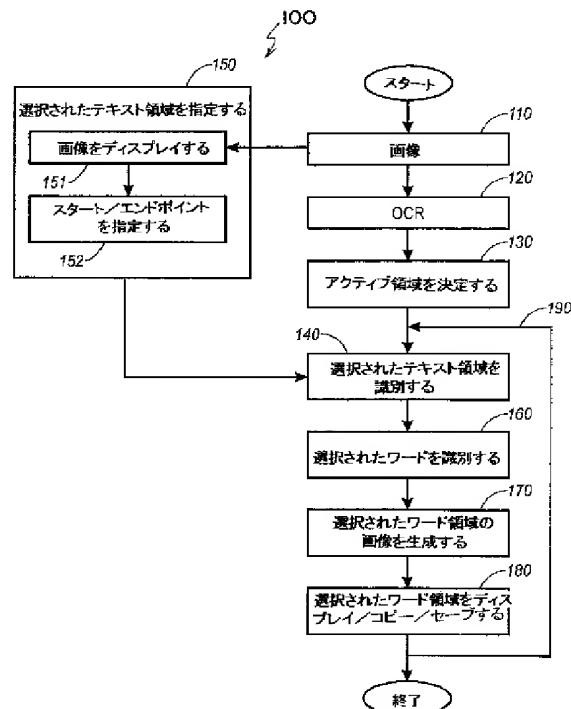
最終頁に続く

## (54)【発明の名称】 テキスト領域選択方法

## (57)【要約】

【課題】 画像からテキスト領域を選択する方法を提供する。

【解決手段】 メモリデバイスから画像内のワード毎の文字及びフォーマット情報にアクセスし、文字及びフォーマット情報に基いた画像内の各ワードに対応付けられたアクティブ領域を決定し、選択されたテキスト領域内の少なくとも一つのワードに対応付けられた少なくとも一つのアクティブ領域に基いて選択されたテキスト領域内の最初と最後のワードを識別し、選択されたテキスト領域内の最初と最後のワードに基いて選択されたテキスト領域内の全てのワードを識別する。



## 【特許請求の範囲】

【請求項1】 (a) メモリデバイスから画像内のワード毎の文字及びフォーマット情報をアクセスするステップと、  
 (b) 前記文字及びフォーマット情報に基いた前記画像内の各ワードに対応付けられたアクティブ領域を決定するステップと、  
 (c) 前記選択されたテキスト領域内の少なくとも一つのワードに対応付けられた少なくとも一つのアクティブ領域に基いて選択されたテキスト領域内の最初と最後のワードを識別するステップと、  
 (d) 前記選択されたテキスト領域内の最初と最後のワードに基いて選択されたテキスト領域内の全てのワードを識別するステップと、  
 を備える画像からテキスト領域を選択する方法。

【請求項2】 前記ステップ (a) に先立って、  
 (e) 画像データからの文字情報を前記画像内のワード毎に生成するステップと、  
 (f) 前記画像内のワード毎に前記フォーマット情報を決定するステップと、  
 (g) 前記画像内のワード毎の前記文字及びフォーマット情報を、前記メモリデバイス内に、記憶するステップと、  
 を備える請求項1に記載の画像からテキスト領域を選択する方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、画像処理に係り、特に、画像からのテキスト領域の選択に関する。

## 【0002】

【従来の技術】 デジタル写真における技術的発展は、コストを削減しながら、デジタルカメラの性能を絶えず向上させており、デジタルカメラは一般のオフィス環境においてドキュメントスキャナとして広く使われるようになった。例えば、ハードコピードキュメントからの画像は、デスクトップの上に位置するカメラによって取得され、更に処理するためにデジタル化され、コンピュータモニタ上にディスプレイされる。このタイプのスキャンは、用紙と電子媒体の間の「スキャン・アズ・ユー・リード ("scan-as-you-read")」インターフェースを促進し、これは、しばしば、「オーバー・ザ・デスク ("over-the-desk")」スキャンと呼ばれる。このようなオーバー・ザ・デスク・スキャンシステムの例としては、"Interactive Copying System (対話型複写システム)"と題された米国特許第5,511,148号において、Wellner (ウェルナ) によって開示されている。

【0003】 オーバー・ザ・デスク・スキャンシステムの従来の接触型スキャンに対する一つの利点は、ハードコピードキュメントをそれらの通常の読み取り場所から取出す(移動させる)必要がなく、これが便利な点である。

これによって、ユーザが、読み取り中に見つけた少量の情報ドキュメントをスキャンすることができるようなカジュアルな形でのスキャンが助長される。例えば、ショートテキストシーケンス(例: 5乃至10ワード)を、カメラに記録し、次に、光学的テキスト認識(OCR)を用いて、コード化されたテキストに変換することができる。

【0004】 選択されたテキストのOCR化(光学的にテキスト認識された)バージョンは、次に、ソフトコピードキュメントへコピーされてもよく、ここで、選択されたテキストはコンピュータシステム上のテキスト編集又はワードプロセッシングアプリケーションによって操作され得る。このタイプのスキャンによって、選択されたテキストをユーザがコンピュータシステムへリタイプする手間が省ける。

## 【0005】

【発明が解決しようとする課題】 選択されたテキスト領域をスキャンする方法では、ユーザは、通常、予想外に多くのテキストやワードを含むテキストのブロック全体を選択しなければならない。ブロック領域全体をスキャンするということは、ユーザが、その選択範囲に含まれるワードを十分にコントロールすることができないということである。

【0006】 選択されたテキスト領域をスキャンする他の方法によれば、ユーザはスキャンしようとするその正確なテキスト領域を選択することができる。恐らく、ユーザは、このタイプの選択が、テキストブロック全体を選択することより、ユーザフレンドリでより便利であることがわかるであろう。正確なワード選択の例としては、"Text/Image Selection From Document Images"

(ドキュメント画像からのテキスト及び/又は画像の選択)と題された公開された欧州特許出願番号98304196.3に記述されている(欧州特許第0881592A2号)。この公報によれば、選択されたテキスト領域は、一画素当り複数のビットを有する画像から抽出され、光学的テキスト認識(OCR)を行う前にデスクew(de-skew)される。

【0007】 Knowlton (ノールトン) への"Method and Apparatus For Automatically Specifying A Portion of Text From a Bitmap Image of The Text" (テキストのビットマップ画像からテキストの一部を自動的に特定する方法及び装置)と題された米国特許第5,517,586号(以下、「ノールトン特許」と称する)には、他の正確なワード選択の例が記載されている。ノールトン特許は、ユーザ指定のスタート及びエンドポイントに応答して、ビットマップ画像の特定された部分の種々の幾何学仕様(例: テキストのラインの横方向の広がり、傾斜、周期性等)を決定する。これらの幾何学的仕様は引き続いてOCRを実行するユーティリティへ渡されてもよい。

【0008】 上述された例は、OCR前にテキスト選択を行なう。OCR前にテキスト選択を行うことの欠点

は、OCR後のテキスト選択と比較して、OCRから導出された情報をテキスト選択を行なうために使用できないことである。これによって応答性が低く（即ち、テキスト選択のプロセスにおいてユーザによる処理時間が増える）、確実性の少ない画像処理システムが生じることになる。

#### 【0009】

【課題を解決するための手段】本発明は、画像からテキスト領域を選択する方法、システム、及びコンピュータプログラム製品を提供する。

【0010】本発明の態様の一つは、(a) メモリデバイスから画像内のワード毎の文字及びフォーマット情報にアクセスするステップと、(b) 前記文字及びフォーマット情報に基いた前記画像内の各ワードに対応付けられたアクティブ領域を決定するステップと、(c) 前記選択されたテキスト領域内の少なくとも一つのワードに対応付けられた少なくとも一つのアクティブ領域に基いて選択されたテキスト領域内の最初と最後のワードを識別するステップと、(d) 前記選択されたテキスト領域内の最初と最後のワードに基いて選択されたテキスト領域内の全てのワードを識別するステップと、を備える、画像からテキスト領域を選択する方法である。

【0011】本発明の態様は、添付図面を参照して、以下に説明することによってより明確に理解されよう。なお、図中、同一の構成要素に関しては、同一の参照符号を付す。

#### 【0012】

【発明の実施の形態】本発明の種々の実施の形態は、一画素当たり複数のビットを有する画像（例：グレースケール、カラー、他）からテキスト部分を選択するための技術に関する。一画素当たり複数のビットを有する画像がビデオカメラや他の同様のデバイス等の画像キャプチャデバイスによって取得され得る。以下に記述されている実施の形態のうちの多くは、ビデオ会議システム、「オーバー・ザ・デスク(over-the-desk)」走査システム、及び他の画像処理システムに特に好適である。

【0013】本発明の好ましい実施の形態においては、ユーザは、スタート及びエンドポイントを指定することによって、テキストを含むディスプレイされた画像の一部を選択する。このディスプレイされた画像は、画像キャプチャデバイス（例えは、カメラ）によって入手された一画素当たり複数のビットを有する画像を表す。その後、ユーザ入力に基いて、スタート及びエンドポイント間の全てのワードを含む選択されたテキスト領域が、光学的文字認識（OCR）中に導出された情報を用いて自動的に決定される。選択されたワード領域が生成され、ディスプレイ及び／又はセーブされ得る。例えは、選択されたワード領域はクリップボードにセーブされ、このクリップボードによって「コピ－・アンド・ペースト」及び／又は「カット・アンド・ペースト」動作を介し

て、アプリケーションプログラムとアプリケーションプログラムの間を、又はアプリケーションプログラム内で、情報が転送される。一般に、クリップボードとは、Microsoft WINDOWS（登録商標）等のオペレーティングシステムによって保持される特殊メモリリソースを称する。本発明の好ましい実施の形態には、特にカメラベースのスキャナに関連して、（ドキュメント画像等）一画素当たり複数のビットを有する画像からの迅速で簡便なテキスト複写方法を提供する。

【0014】本発明の種々の実施の形態によるワード選択技術はスクリュー又はワープされた画像上で実行することもできる。

【0015】選択されたワード領域を決定するためにOCRから導出された情報を使用することにはいくつかの利点がある。本発明の好ましい実施の形態によれば、OCR（光学的文字認識）の結果には、ドキュメントセグメンテーション情報（例えは、ドキュメント画像の種々のテキスト、画像（ピクチャ）及びラインセグメント部分の識別に関する情報）、テキスト情報（例：ASCII（アスキー）テキスト）、及びフォーマッティング情報（例：ワード境界ボックス、テキストライン、及びテキスト領域）が含まれる。

【0016】OCR後ではなくOCR前にテキスト選択を実行するテキスト選択方法と比較して、本発明の実施の形態は、選択されたテキスト領域のフィードバックに對して一層増大した応答性を有する。これは、選択されたテキスト領域を指定するユーザへ応答するのではなく、ユーザ入力に先立って画像処理及びセグメンテーションの大部分を実行することによって達成される。この実施の形態においては、選択されたテキスト領域が識別されると、最小限の演算で済ませることができる。従つて、ディスプレイへフィードバックされる選択されたテキスト領域を得る前に、ユーザが待たなければならない時間は減少される。これによって、本発明の実施の形態のテキスト選択技術の有用性と生産性は高まる。

【0017】更に、セグメンテーション（テキスト対ノンテキスト）は、ユーザによって定義付けられたテキスト領域に基くのではなく画像全体に依存するので、本実施の形態のテキスト選択技術の精度及び反復性は、知られているテキスト選択方法よりも良くなっている。（画像全体と比較して）選択されたテキスト領域から得られるセグメンテーションの結果は、選択されるワード数又はワード選択時にユーザが最初にクリックする場所によって変化し得ることに注目されたい。対照的に、画像全体に対するOCRから（又はOCRに先立って）得られるセグメンテーションの結果を用いる時、与えられるワードの選択は、選択されたかもしれない他のワードとは無関係に一定している。さらに、OCRからのセグメンテーションの結果によって画像内の不要な成分（例：ノンテキストアイテム）を除去くことができる。スケール

もまた、選択されるテキスト領域より、画像全体を使用する時に、より確実に検出され得る。さらに、テキストの劣化を記述するとともにテキスト領域の文法的なコンテキストを記述するパラメータは、OCRを確実にするには重要である。また、小さな画像領域からこれらのパラメータを推定することは難しい。従って、OCRが使用される場合、ユーザーによって指定されるほんの小さなエリアよりも大きな領域を処理するほうが有利である。

【0018】更に、本実施の形態のテキスト選択方法はOCRの結果を利用してより広い範囲のコンテキストへと選択の幅を広げることができる。例えば、読み出し順を言語（例：中国語、ペルシア語、その他）に適応させることができる。テキスト選択の適応における他のOCRの結果の使用例としては、チャート、図、及び表におけるテキストの向きの検出である。また、字体（例えば、アセンダ、デセンダ、その他）や句読点（例えば、ハイフン、その他）によって文字間隔や行間隔を変更することもできる。

【0019】図1は、本発明の一つの実施の形態による、画像からテキスト領域を選択するための論理フローチャート100を示す。一画素当たり複数のビットを有する画像は、ボックス110に示されるように、スキャンデバイス、ビデオカメラ等の画像キャプチャデバイスから得られる。ビデオカメラから画像を取込む時、ユーザーが選択したテキスト領域のスタートポイント（又は他のポイント）を指定した上で、ユーザーが「フリーズ」ボタン、「選択モード」ボタンを選択したり、カメラのモーションを停止させるための知られている他の方法によって、画像を「フリーズ」させることができる。

【0020】得られた画像はしばしば、OCRに合った画像にするために、種々の公知の画像形成技術（例：セグメンテーション、2値化等）によって、処理する必要がある。例えば、OCRアルゴリズムは、入力画像としてバイナリ（2値）画像を必要とすることもある。一般に、光学的文書認識（OCR）とは、ビットマップされた画像の表示を、対応付けられる文字やシンボルに変換することを称する。用語「ビットマップ画像」とは、各画素が一つ以上のビットによって表される任意の画像を称する。

【0021】画像にはテキストが含まれ、また、ノンテキストアイテム（例えば、絵）も含まれる。用語「テキスト」は使用される特定のOCRエンジンによって認識されることが可能な、文字、シンボル、及び／又はオブジェクトを称し、「ノンテキスト」は画像内の全ての他のアイテムを称する。従って、テキスト領域は、一般に、英語又は他の言語に使用されるアルファベット、数字、オブジェクト、及び／又はシンボルを含む。図2は、テキストを含む画像200の例を示す。概して、ワードは一つ以上の認識されたテキスト文字によって形成される。

【0022】ボックス120において、画像上で実行されるOCRは、当業者に良く知られている多数のOCRエンジンによって実行されてもよい。OCR中、画像のビットマップ表示は、解釈されて、コード化されたテキストへ変換される。このテキスト情報は、フォーマッティング情報（例えば、各ワードに対応付けられる境界ボックス、テキスト領域、及びテキストライン）と共に、ファイルに記憶されてもよい。

【0023】本発明の好ましい実施の形態に関して、Scansoft, Inc.（スキャンソフト社）によって製造されたTextbridge（テキストブリッジ）OCRエンジンは、画像上でOCRを実行するために使用される。図3は、OCRエンジン120によって生成された結果を表すプロック図を示す。OCRの結果は、本発明の好ましい実施の形態の境界ボックス情報310、テキスト情報320、テキストライン情報330、及びテキスト領域情報340を含む。図3において、OCRエンジン120は、各ワード毎に境界ボックス、各ワードに対応付けられるテキスト領域及びテキストライン、及び各ワード内で認識される文字を定義付ける情報を生成する。好ましい実施の形態によれば、テキストモジュール320は、情報交換用米国標準コード（アスキー（ASCII））と呼ばれるコード体系を使用し、このコード体系は、文字、数字、句読点、及び他の文字に数値を割当てるために使用される。他の実施の形態において、ASCII以外のコード体系を使用してもよい。Textbridge OCRエンジンは、OCRの結果を、テキスト及びフォーマッティング情報の両方を含むXDOCファイル内にセーブし、このXDOCファイルをクリップボードに記憶する。

【0024】境界ボックスは、1ワード（語）当りの文字全体を含む、画像座標内の最小のボックスを称する。図4は、画像200内のテキストに対応付けられる境界ボックスを示す。具体的には、各ワードを囲む矩形が境界ボックスを表す。

【0025】テキスト領域は、グループ化された画像内のテキストの領域に関する。要するに、各テキスト領域は、他のいかなるブロックにも接続されないテキストブロックを表す。例えば、テキストの各コラムは、新聞の記事の各コラムが分離したテキスト領域として分類されるようにテキストの領域を表すこともある。テキストの領域はしばしば矩形であり、テキスト領域境界によって定義付けられる。更に、複雑なレイアウトにおいてはテキストの領域が互いにオーバーラップすることもある。図5は、画像200のテキスト領域に対応付けられるテキスト領域境界510を示す。好ましい実施の形態のOCRエンジンは、各テキスト領域に固有の領域識別番号を割当て、各ワードに対応付けられるテキスト領域を決定する。

【0026】各テキストラインは、同一テキスト領域内

にあり、互いに隣り合って水平に現れる、多数の言語のワードを含む。各テキストラインは、上部テキストライン境界及び下部テキストライン境界によって定義付けられる。図5に示されるテキストライン境界540は、画像200の（ワードではなく一続きの境界ボックスによって示される）第2のテキストラインのための下部テキストライン境界を示す。好ましい実施の形態におけるOCRエンジンテキスト領域内の各テキストラインに対してライン番号を割当て、各ワードに対応付けられるテキストラインを決定する。

【0027】他の実施の形態によれば、OCRエンジンが、これらのテキスト領域及びテキストラインを定義付ける情報を生成しないこともある。このような実施の形態に対しては、テキスト領域及びテキストラインに関する情報は、境界ボックス情報に基いて、演算されてもよい。例えば、テキスト領域は、上下部に隣接する距離内に置かれたワード全体をクラスタ化することによって探索され得る。テキスト領域が識別されると、境界ボックスの中心を垂直軸へ投影し、これらのポイントをクラスタ化することによって、テキストラインが探索される。各クラスタセンタがライン番号を割当てられる。ワードは、所与のクラスタセンタに交差する水平線に最も近くなったとき、所与のラインに属する。

【0028】図1に戻って参照する。OCRの結果が得られると、ボックス130に示されるように、複数のアクティブ領域が決定され得る。本発明の一つの実施の形態によれば、図2に示されるように、一つのアクティブ領域が、画像200内の各ワードのような、画像内のワード毎に定義付けられる。各アクティブ領域は、複数の画像座標ポイントを表し、これによって、画像座標ポイント（例えば、スタートポイント、エンドポイント、又は他のポイント）が与えられると、このポイントに最も近いワードが探索され得る。好ましい実施の形態においては、このポイントは、ユーザがマウスやキーボードなどのカーソルコントロールデバイスを使って、指示すこともできる。

【0029】図5は、第1のテキストライン内のワード全体、及び第2のテキストライン内の最初の二つのワードのアクティブ領域を示す。便宜上、画像200内のワードのうちのいくつかのワードに対するアクティブ領域のみが示されている。しかしながら、画像内の全てのワードに対してアクティブ領域が一般的にはこのように決定されることに注目されたい。

【0030】アクティブ領域は、画像内のワードがワード上又はワード近くのポイントに基いて識別されるのを可能とする。例えば、選択されたテキスト領域のスタートポイントとエンドポイント（図10に示したスタートポイント1010及びエンドポイント1020）が与えられると、スタートポイントとエンドポイントに対応付けられるアクティブ領域が探索されてこれらのポイント

に最も近いワードを識別することができる。従って、選択されたテキスト領域を定義付けるために、指定されたポイント（ユーザによって指定されるか否か別として）を使用することにより、本発明の実施の形態のワード選択技術は、OCRの結果を選択されたテキスト領域にマッチさせることができる。これによって、適切なワードを他のファイルへコピーすることができる。

【0031】一般に、アクティブ領域は、OCR段階中に生成された情報を使って、ワード間そしてテキストライン間の部（ディビジョン）を探索することによって決定される。具体的には、アクティブ領域の垂直境界はワードディビジョンを探索することによって決定され、アクティブ領域の水平境界はラインディビジョンを探索することによって決定される。

【0032】図6は、本発明の一つの実施の形態によるアクティブ領域の垂直境界を決定する論理フローチャート600を示す。ボックス610において、隣接する境界ボックスの垂直境界間の中間点が決定される。図8は、四つの境界ボックス801乃至804を有するワードラインを説明する。中間点810、820、及び830は、隣接する境界ボックスの垂直境界間の中間点の例を示す。

【0033】二つの隣接するワードを有するワード（「センターワード」とも呼ばれる）毎に、左側の中間点から右側の中間点へ延びる領域はアクティブ領域として指定される（ボックス620）。このように、左側中間点と右側中間点はセンターワードに対するアクティブ領域の垂直境界を表す。

【0034】ボックス630に示されるように、一つの隣接するワードを有するワード毎に、その右側中間点からその左側テキスト領域境界へ又はその左側中間点からその右側テキスト領域境界へ延びる領域がアクティブ領域として指定される。例えば、図8において、境界ボックス801の垂直アクティブ領域境界は垂直テキスト領域境界840から中間点810へ延びると共に、境界ボックス804の垂直アクティブ領域境界は垂直テキスト領域境界850から中間点830へ延びる。

【0035】テキストライン内にワードが一つだけ残った場合、アクティブ領域は、テキスト領域の二つの垂直境界間に延びる。

【0036】図7は、本発明の実施の形態によるアクティブ領域の水平境界を決定するための論理フローチャート700を示す。ボックス710においては、各テキストラインの上部及び下部境界が決定される。本発明の一つの実施の形態に関しては、テキストラインの上部境界はその特定のテキストライン内の任意の境界ボックスの最上ポイントとして定義付けられ、同様に、テキストラインの下部境界はその特定のテキストラインにおける任意の境界ボックスの最下ポイントとして定義付けられる。

【0037】ボックス720において、隣接するテキストラインの上部及び下部の境界間の中間点が計算される。例えば、図9は、三つのテキストライン901乃至903を示す。中間点910はテキストライン901とテキストライン902の間の中間点を表し、中間点920はテキストライン902とテキストライン903の間の中間点を表す。

【0038】図7に戻って参照すると、ボックス730とボックス740において、アクティブ領域の水平境界が決定される。ボックス730においては、二つの隣接するテキストラインを有するテキストライン（「センター・テキストライン」とも呼ばれる）に対してアクティブ領域の水平境界が決定される。センター・テキストラインのアクティブ領域に対する水平境界は、上部中間点から下部中間点へ延びる。例えば、図9に示される中間点910から中間点920までは、テキストライン902に対する水平境界を表す。特定のテキストライン内の全てのアクティブ領域は同じ水平のアクティブ領域境界を有することに注目されたい。

【0039】図7に戻って参照すると、ボックス740において、アクティブ領域の水平境界が、たった一つの隣接テキストライン（「サイド・テキストライン」と呼ばれる）を有するテキストラインに対して決定される。サイド・テキストラインのアクティブ領域に対する水平境界は、テキスト領域境界から中間点へ延びる。例えば、図9に示されるように、テキストライン901内のアクティブ領域に対する水平境界は、水平テキスト領域境界930から中間点910へ延び、テキストライン903内のアクティブ領域に対する水平境界は、水平テキスト領域境界940から中間点920へ延びる。

【0040】画像内にテキストラインがたった一つ残った場合、アクティブ領域の水平境界はそれに対応付けられるテキスト領域の境界から延びる。

【0041】画像内の各ワードは対応付けられるアクティブ領域を有する。本発明の一つの実施の形態によれば、アクティブ領域の境界は、論理フローチャート600及び700に従って決定される。

【0042】本発明の他の実施の形態によれば、ライン及びワードのディビジョン（即ち、水平及び垂直境界決定）がコンテキストにより適合した方法で実行され得る。例えば、ライン・ディビジョンは、文字が、アセンダかデセンダかを考慮することによって決定され、ワード・ディビジョンは、ハイフン付きのワードか否かを考慮する。アセンダは、英字の小文字の英字の本体より上に出る部分をいい、デセンダは、基線の下にはみ出ている部分をいう。ライン・ディビジョンは、文章の始まりに一般に発生する文字見出しの存在をも考慮に入れる。

【0043】本発明のさらに他の実施の形態によれば、アクティブ領域は他の何らかの方法で配列されてもよい。例えば、選択されたポイントに最も近いワードが選

択されるのではなく、選択されたポイントの左側のワードが選択される。これは、アクティブ領域を左側の境界ボックスと位置合わせすることによって達成され得る。

【0044】OCRの結果（と異なる情報）はビットマップ画像と共にファイルやメモリに記憶されてもよく、これによって、ワードツーワードの選択は、OCRの結果を再演算せずに後から実行することができることに注目されたい。

【0045】対応付けられるアクティブ領域が決定されると、図1のボックス140に示されるように、選択されたテキスト領域を、ユーザにより、半自動的、又は自動的のいずれかで指示した上で、選択されたテキスト領域が識別され得る。選択されたテキスト領域は、一つ以上のポイント（例：スタートポイント及びエンドポイント）を指定することによって指定され得る。

【0046】本発明の好ましい実施の形態によれば、画像（又は画像の任意の表示）がディスプレイされ（ボックス151）、ユーザは、マウス、キーボード等のカーソルコントロールデバイスを用いて、選択されたテキスト領域を指定するためにスタート及びエンドポイントを指定する（ボックス152）。指定されたポイントの各々は、アクティブ領域内にあり、対応付けられるワードを有する。スタートポイントによって識別されるワードは選択されたテキスト領域内の最初のワードとされ、エンドポイントによって識別されるワードは選択されたテキスト領域内の最後のワードとされる。

【0047】画像は、画像キャプチャデバイスによってキャプチャされたオリジナル画像、オリジナル画像のエンハンスされた画像、オリジナル画像の2値化されたバージョン、オリジナル画像に最も類似するOCRテキストの表示バージョン、領域を示したり、任意の他の有用な情報によって画像を増補するバージョン等の種々の形態でユーザへディスプレイされる。

【0048】スタートポイント及びエンドポイントを指定する際、ユーザは、マウスを使って、スタートポイントをクリックし（スタートポイントを指定し）、次に、（エンドポイントを指定するために）エンドポイントまでマウスポインタをドラッグし、その時点で、ユーザはマウスボタンを離してもよい。ユーザは、マウスを使って、スタートポイントをクリックし、次に、エンドポイントをクリックしてもよい。

【0049】図10は、スタートポイント1010とエンドポイント1020を示す。スタートポイント1010は、ワード"every"に対応付けられるアクティブ領域内にあり、エンドポイント1020は、ワード"reports"に対応付けられるアクティブ領域内にある。スタートポイント1010とエンドポイント1020に基いて、これらのスタートポイントとエンドポイントに対応付けられたアクティブ領域が識別される。本発明の一つの実施の形態においては、指定されたポイント（例えは、ス

タートポイントとエンドポイント)のディスプレイ座標をその対応するアクティブ領域にマッピングするためにロックアップテーブル(LUT)を使用してもよい。各アクティブ領域を複数の座標ポイントによって表してもよい。これらのアクティブ領域に基いて、ワード"every"は、選択されたテキスト領域内の最初のワードとして識別され、ワード"reports"は、選択されたテキスト領域内の最後のワードとして識別される。

【0050】本発明の他の実施の形態においては、カーソルコントロールデバイスや他の選択用デバイスを用いてユーザーにスタートポイントとエンドポイントを選択させる代わりに、これらのポイントを半自動的又は自動的に決定してもよい。スタートポイントとエンドポイントをユーザーの入力によらずに自動的に選択してもよいし、ユーザーにスタートポイント又は他のポイントのみを選択させるように、半自動的に選択してもよい。半自動選択の例としては、スタートポイントをユーザーに選択させ、次に、残りのセンテンスが自動的に選択されること、又は、ユーザーにパラグラフ内の任意の場所をダブルクリックさせて、次に、パラグラフ全体を選択することを含む。任意の選択の例としては、いくつかのクラス(例:名前、eメールアドレス、ハイパーリンク等、又は入力されたワードに適合する全てのワードを選択すること)に入っている全てのワード又はフレーズを選択することを含む。

【0051】他の実施の形態においては、ユーザーは、キーボード上で他のキーを押すことによって複数のテキスト選択を実行してもよい(例:CTRL又はshiftボタン)。また、同様に、テキスト選択を結合したり、これらを併合したり、現存する(複数の)ワードから(複数の)ワードを除くことも可能である。さらに、テキスト選択は、より複雑なレイアウト(例:セル、行、列を含む表)やフォーマット情報(例:表におけるセル、行、列を選択する為に必要とされる)を配慮することもできる。

【0052】スタートポイントとエンドポイントが(ユーザーにより、又は半自動的、自動的のいずれかで)指定されると、選択されたテキスト領域内の最初と最後のワードが識別される。図1のボックス160に示されるように、選択されたテキスト領域内の最初と最後のワードを使って、選択されたテキスト領域内に入っている他の全てのワードが識別される。

【0053】本発明の一つの実施の形態においては、選択されたテキスト領域内の全てのワードが図14に示された論理フローチャート1400によって識別される。ボックス1410において、最初のワードと同じテキストライン上のこの最初のワードを含み、(この最初のテキストワードの)右側にある全てのワードを選択する。ボックス1420において、最後のワードと同じテキストライン上のこの最後のワードを含み、(この最後のワードの)左側にある全てのワードを選択する。ボックス1430においては、最初のワードと最後のワードの間の全てのライン上の全てのワードを選択する。

【0054】本発明の一つの実施の形態においては、図11は、最初と最後のワードが識別されたときに選択されたテキスト領域内の全てのワードを識別する際の例を示す。ワード1130は最初のワードを示し、ワード1140は最後のワードを示す。ライン1上のワード1130の右側の全てのワード(即ち、ワード1131から1135まで)が選択される。ライン3上のワード1140の左側の全てのワード(即ち、ワード1141から1143まで)が選択される。ライン2上の全てのワード(即ち、ワード1150から1158)が選択される。

【0055】図1に戻って参照すると、ボックス170において、選択されたワード領域の画像が生成される。本発明の一つの実施の形態において、この選択されたワード領域は、図15に示される論理フローチャート1500に沿って生成される。

【0056】図1に戻って参照すると、ボックス170において、選択されたワード領域の画像が生成される。本発明の一つの実施の形態において、この選択されたワード領域は、図15に示される論理フローチャート1500に沿って生成される。各選択矩形の上下の境界は、同一テキストライン内のアクティブ領域の上下境界の矩形と同一である。選択矩形の左右の境界は、最初に選択されたワード(例:ワード1210)境界ボックスの左側と最後に選択されたワード(例:ワード1220)境界ボックスの右側によって決定される(図12参照)。

【0057】ボックス1520においては、選択矩形は一つ以上のポリゴン(多角形)に併合される。例えば、選択矩形1230、1240、及び1250は、図13に示されるポリゴン1310に併合される。ポリゴンは選択されたワード領域を表す。

【0058】選択されたワード領域は、生成されると、図19に示されるように、ディスプレイされる。本発明の好ましい実施の形態によれば、選択されたワード領域は自動的に現れ、ユーザーがマウスポインタを動かすと再度描画される。選択されたワード領域をディスプレイするとき、ポリゴン領域は、画像のいくつかの変換に応じて、ディスプレイされる。例えば、ポリゴン領域に対応するレンダリングされたバージョンのOCR化されたテキストがディスプレイされてもよい。或いは、一画素当たり複数のビットを有する画像(グレースケール画像又はカラー画像)、バイナリ画像等がディスプレイされる。本発明の好ましい実施の形態においては、選択されたワード領域がビデオ反転領域としてディスプレイされる。

【0059】選択されたワード領域をディスプレイすることに加えて、選択されたワード領域がクリップボード内にセーブされ、引き続いて、種々のワードプロセッシ

ングアプリケーション等のアプリケーションプログラム（「カット・アンド・ペースト」及び／又は「コピー・アンド・ペースト」）へ転送される。Microsoft WINDOWS（登録商標）のクリップボードへコピーするとき、レンダリングされたバージョンのASCIIテキストが使用される。更に、選択されたワード領域がメモリデバイス内に記憶されてもよい。

【0060】ある画像に対するアクティブ領域を決定した後は、一つ以上のテキスト選択を一つの画像に対して実行することができ、その際、境界ボックス情報、テキストライン情報、テキスト領域情報、アクティブ領域の位置及び範囲、及び画像のワープを再度演算する必要がないことに注目されたい。他の選択は、矢印190によって示されるように、ボックス140、160、170、及び180で実行される機能を反復することによって行われる。

【0061】本発明の他の実施の形態においては、ワード選択は、スキー（図17参照）されたり、貞の湾曲、遠近効果（図16参照）や樽状歪みによってワープされた画像上で実行される。ワープされたワード／テキストの選択を実行する機能は、画像キャプチャデバイスとしてビデオカメラを使用するときに特に有用である。透視歪み、樽状歪み、及びスキーは、「オーバー・ザ・デスク」スキャンシステムによって得られるカメラ画像においてしばしばみられる。

【0062】図18は、ワープされた画像上でテキスト選択を実行するための論理フローチャート1800を示す。ワープされた画像上でワード又はテキスト選択を実行するとき、ボックス1810、1820、1830、及び1840に示されている更なるステップが実行されてもよい。図18におけるボックスのうちのいくつかが、図1に示されたボックスに対応しており、従って、これらのボックスについての同様の説明がここでは省略されることを理解されたい。

【0063】OCR（ボックス120）に先立って、画像内のワープが検出されてもよい（判定ボックス1810）。画像内のワープが検出された場合、ボックス1820内で推定されるワープ関数を用いることによって、OCRに先立って、ワープが補正されてもよい（ボックス1830）。

【0064】本発明の一つの実施の形態において、画像のワープは、デワープ（dewarp）（ワープ補正）された画像D内のポイントqを観察画像O内のポイントpにマッピングする関数Wによって記述され得る。観察画像Oは、カメラによって記録されたり、他の画像キャプチャデバイスによって取得された画像を表すこともある。通常、ワープは、一セットのパラメータによってパラメータ化され、 $p = W(q, a)$ となる。例えば、スキーされたページを記述するには、画像の水平方向に対してテキストラインを回転させるために、aは一つだけのパ

ラメータから成るように選択され得る。他の実施例によれば、平らなページに対して遠近記述するには、カメラの光軸に直交する三つの軸に対するページの回転を記述するために、aは三つのパラメータから成るように選択され得る。もっと一般的な透視的状況では、光軸、光学的センタ、及びカメラの焦点長さに沿ったページの距離を記述するために更なるパラメータが必要とされる。ポイントpでのグレイ値（又はカラー値）O(p)で観察画像Oをデワープするには、グレイ値（又はカラー値）D(q) = O(W(q, a))で、画像Dを構築する。

【0065】Oの画素の格子上にW(q, a)が現れない一般的な状況において、W(q, a)に最も近い格子点を取る。或いは、W(q, a)の近傍に格子点のグレイ値を補間する。

【0066】ボックス1820においてワープ関数のパラメータが推定される。ワープ関数は、入力画像の座標を出力画像の座標にマッピングし、これによって、入力画像ポイント(x, y)とすると、出力画像(x', y')における対応する座標は、 $x' = a(x, y)$ 及び $y' = b(x, y)$ となる。本発明の一つの実施の形態において、ワープ画像（即ち、スキーによる）は、角度 $\alpha$ （単一パラメータa）分、画像を回転させることによってデワープされ、ポイントp(x, y)の座標は、以下のように、ポイントq(x', y')にマッピングされる。

$$x' = x \cos(\alpha) - y \sin(\alpha)$$

$$y' = x \sin(\alpha) + y \cos(\alpha)$$

【0068】他の実施の形態では、当業者によく知られた他のワープ関数も使用することができる。

【0069】ボックス1820で推定されたワープ関数は、ボックス1840で選択されたワード領域に最終的に適用される。選択されたワード領域はデワープされた画像から選択されたことに注目されたい。この選択が、スタートポイントとエンドポイントが選択されたオリジナルのワープ画像に合致することを確実とするために、選択されたワープ領域はボックス1820において生成されたワープ関数を用いることによってワープされる。

【0070】一方、ボックス1810において画像内でワープが検出されない場合、ボックス1830はバイパスされ、ここでは、画像をデワープせずに画像上でOCRを実行しても良い。

【0071】図20は、上述された実施の形態によるテキスト選択技術を実行するために使用される画像処理システム2000の例を示す。本発明の一つの実施の形態において、画像処理システム2000は、ソフトウェアコンポーネント2004及びハードウェアコンポーネント2002を有するコンピュータシステムであっても良い。コンピュータシステムは、パーソナルコンピュータ、ワークステーション等であってもよい。好ましい実施の形態においては、本発明は、Windows（登録商標）

NTを実行するIBM互換性のPCのC++において実施される。

【0072】ハードウェアコンポーネント2002は、プロセッsingユニット（即ち、CPU）2006、ランダムアクセスメモリ（RAM）2008、リードオンリーメモリ（ROM）2010、ユーザ入力／出力（I/O）2021、及びネットワークI/O2014を有する。ユーザ入力／出力（I/O）2021は、キーボード2016、カーソルコントロールデバイス2017（例、ポインティングスティック、マウス、他）、ディスプレイ2012、フロッピーディスクドライブ2031、ハードディスクドライブ2032等の種々の入力及び出力デバイスに連結され得る。

【0073】RAM2008は、一般的な記憶領域として及びスクラッチパッドメモリとして、CPUによって使用され、ディスプレイ2012上のディスプレイ（表示）のためのダウンロードされた画像データ（例：一画素当り複数のビットを有する画像データ、バイナリ画像データ、レンダリングされたバージョンのテキスト画像データ等）を記憶するためにも使用され得る。更に、RAM2008内のクリップボードが指定されてもよい。ROM2010はCPU2006によって実行される命令を記憶する為に使用され得る。

【0074】ディスプレイ2012は、画像処理システム2000内のCPU2006又は他のコンポーネントによって提供される画像データをディスプレイする出力デバイスである。本実施の形態において、ディスプレイ2012は、当業者に公知である、スクリーン上に画像をディスプレイする画素の矩形のグリッドを表す、ラスタデバイスである。即ち、ビットマップは、ディスプレイ2012に入力することができ、ビットマップのビットはピクセル（画素）としてディスプレイされ得る。入力されたピクセルマップは、良く知られているように、ディスプレイ2012上に直接表示されるか、又は画像処理システム2000のコンポーネントが、ページ記述ファイルからのコードや他の画像記述を、まず、ビットマップにレンダリングして、これらのビットマップを、ディスプレイ2012上にディスプレイされるように転送する。CRT（カソードレイチューブ）、LCD（液晶ディスプレイ）等のディスプレイは、本発明に適している。

【0075】更に、ユーザI/O2021は、カメラシステム、スキャンデバイス（例：フラットベッド、又はハンドヘルドスキャナ）などの画像キャプチャシステム2030、又は他のハードコピードキュメントの画像の取得用デバイスに連結されても良い。

【0076】更に、ユーザI/O2021は、画像データを記憶するためのフロッピー（登録商標）ディスクドライブ2031及び／又はハードディスクドライブ2032に連結されてもよい。ノンボラタイル（不揮発性）

メモリ（例：フラッシュメモリ）、PCデータカード、その他のような他の記憶デバイスを、画像処理システム2000によって使用されるデータを記憶するために使用することもできる。

【0077】ネットワークI/O2014は、LAN（ローカルエリアネットワーク）、WAN（広域エリアネットワーク）、インターネット他への通信のゲートウェイを提供する。ネットワークI/O2014は、一つ以上のコンピュータシステム又は周辺デバイス（例えば、画像キャプチャシステム2030）に連結されたネットワーク2015へ又は同ネットワークからのデータの送受のために使用される。

【0078】ソフトウェアコンポーネント2004は、オペレーティングシステムソフトウェア2016、OCRモジュール2020、アクティブ領域モジュール2018、ワードモジュール2013、及び選択されたワード領域モジュール2022を有する。オペレーティングシステムソフトウェア2016は、MS-DOS、マッキントッシュOS、OS/2、WINDOWS（登録商標）、WINDOWS（登録商標）NT、又はUNIX（登録商標）オペレーティングシステム又は他の公知のオペレーティングシステムを表してもよい。OCRモジュール2020は、ピクセルマップ画像表示を、ワード位置情報も含む変換テキスト画像情報へ変換する。アクティブ領域モジュール2018は、テキスト画像内の各ワードに対応するアクティブ領域を決定する。選択されたワード領域モジュール2022は、指定されたスタートポイント及び／又はエンドポイントをOCR化された結果にマッピングすることによって選択されたワード領域を生成する。

【0079】

【発明の効果】本発明のテキスト領域選択方法は、ビデオ会議システム、「オーバー・ザ・デスク（over-the-desk）」スキャンシステム、及び他の画像処理システムに特に好適である。

【図面の簡単な説明】

【図1】本発明の一つの実施の形態による画像からのワード選択のための論理フローチャートである。

【図2】テキストを含む画像の一例を示す図である。

【図3】本発明の一つの実施の形態によるOCR（光学的テキスト認識）の結果を示すブロック図である。

【図4】ワード境界ボックスを有する画像の一例を示す図である。

【図5】ワード位置情報と、画像に対応付けられたアクティブ領域の例を示す図である。

【図6】本発明の一つの実施の形態によるアクティブ領域の垂直境界を決定する論理フローチャートである。

【図7】本発明の一つの実施の形態によるアクティブ領域の水平境界を決定する論理フローチャートである。

【図8】アクティブ領域の垂直境界の一例を示す図である。

【図9】アクティブ領域の水平境界の一例を示す図である。

【図10】アクティブ領域の水平境界の一例を示す図である。

【図11】選択されたテキスト領域内の全てのワードを識別する際の例を示す図である。

【図12】選択されたワード領域を生成する矩形を定義付ける際の例を示す図である。

【図13】選択されたワード領域を生成するポリゴンを定義付ける際の例を示す図である。

【図14】本発明の一つの実施の形態による選択されたテキスト領域内の全てのワードを識別するための論理フローチャートである。

【図15】本発明の一つの実施の形態による選択されたワード領域を生成するための論理フローチャートである。

【図16】遠近ワープを有する画像の例を示す図である。

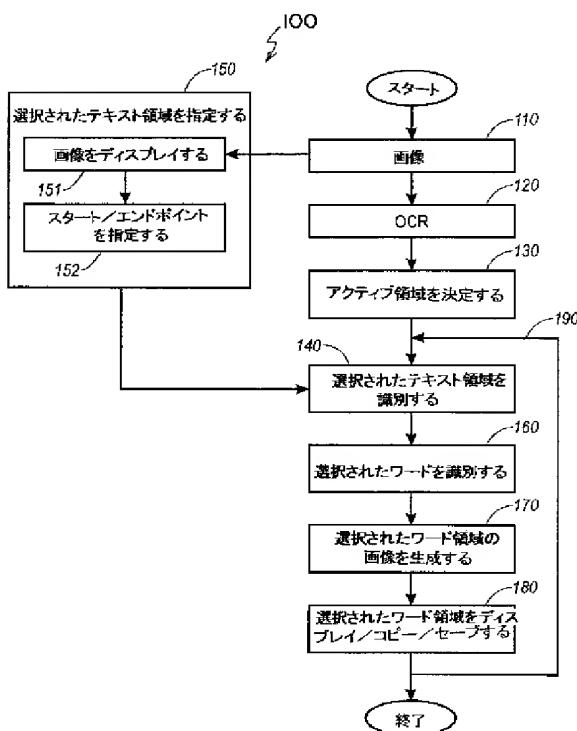
【図17】スキーを有する画像の例を示す図である。

【図18】本発明の一つの実施の形態による、ワープした画像からワードを選択するための論理フローチャートである。

【図19】本発明の一つの実施の形態による、ユーザへのテキスト選択のフィードバックを示す図である。

【図20】本発明の一つの実施の形態による、画像処理システムを示すブロック図である。

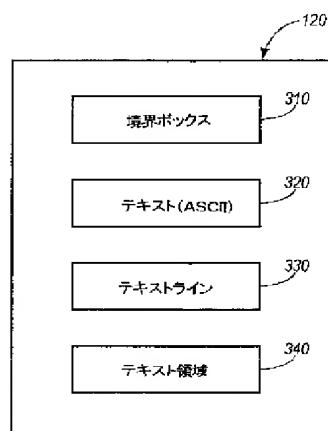
【図1】



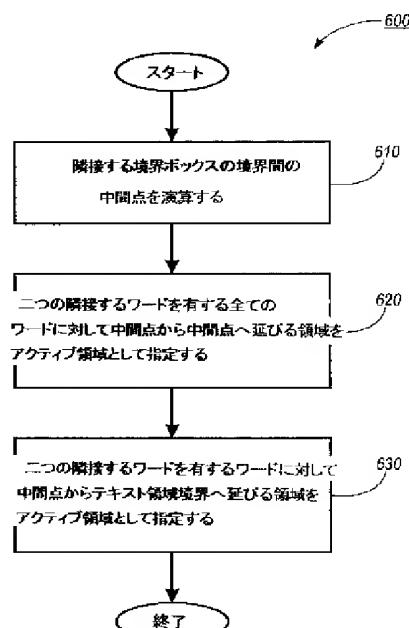
【図2】

People from virtually every branch of professional work author electronic documents while referring to paper documents such as books, articles and reports. Despite efforts to develop paperless authoring environments, there is little evidence that the use of paper

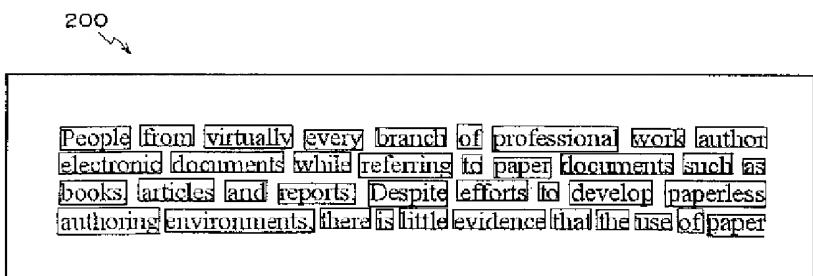
【図3】



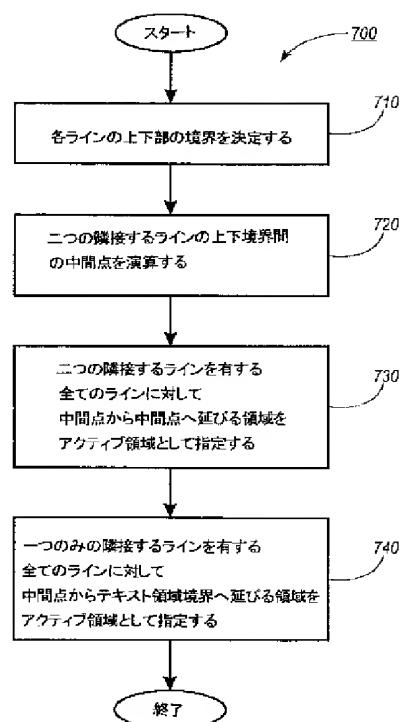
【図6】



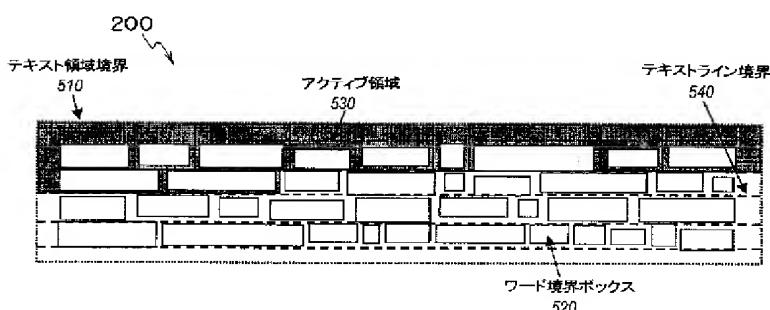
【図4】



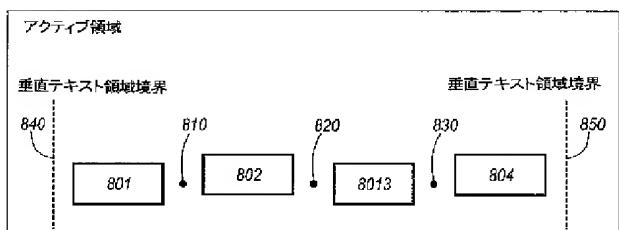
【図7】



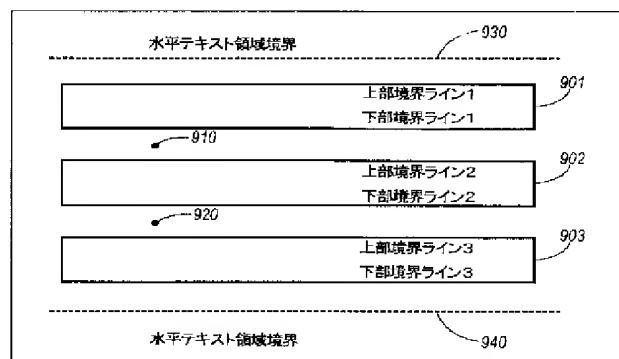
【図5】



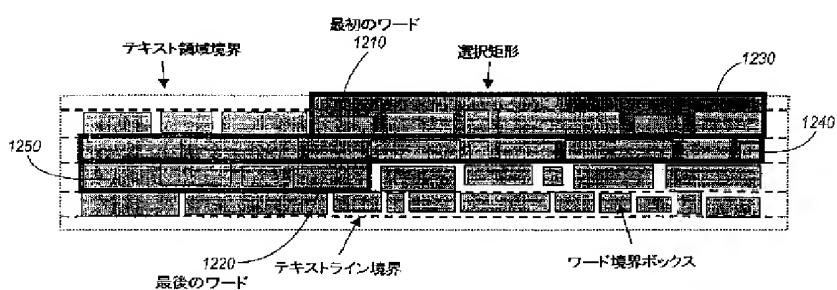
【図8】



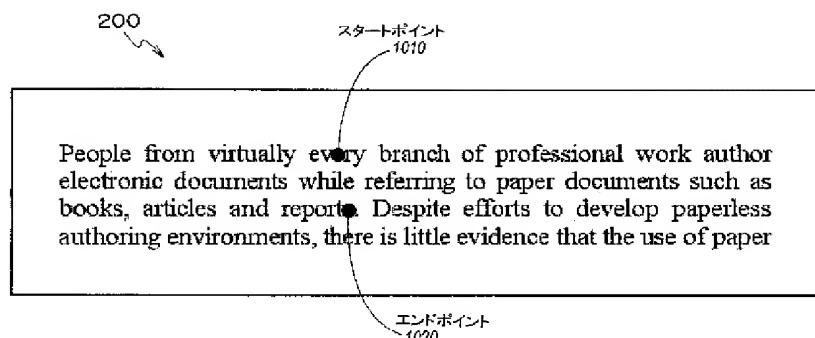
【図9】



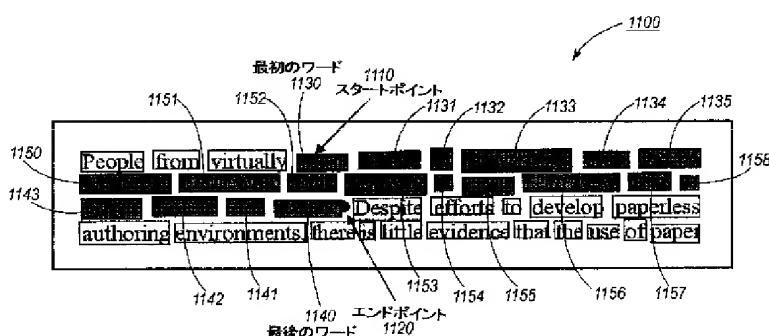
【図12】



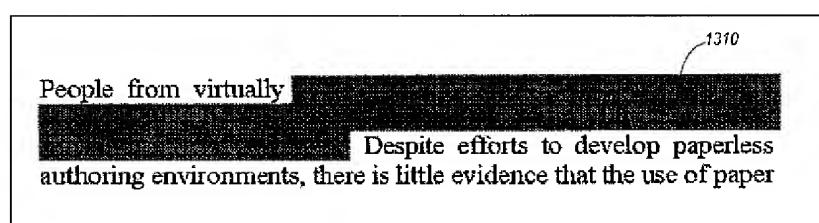
### 【図10】



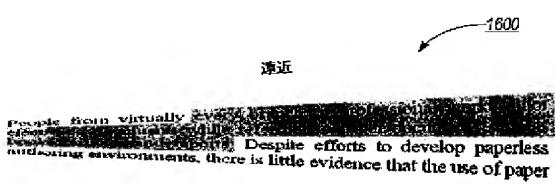
【図11】



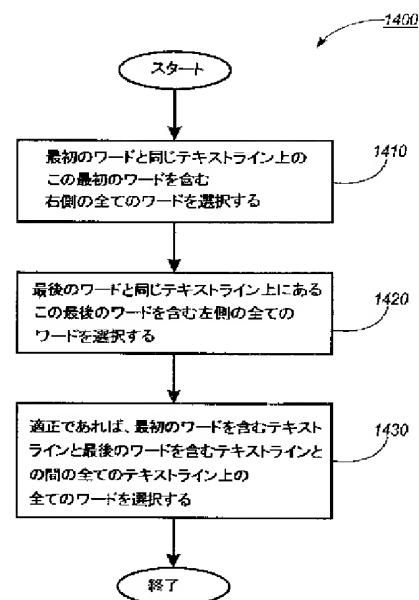
【図1-3】



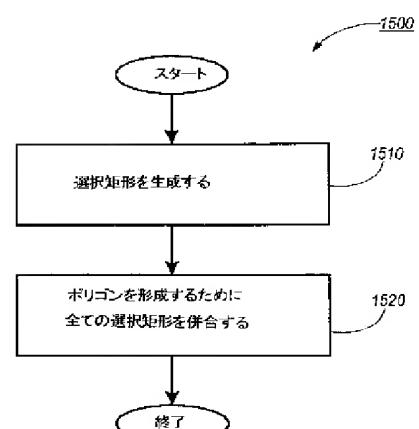
【図16】



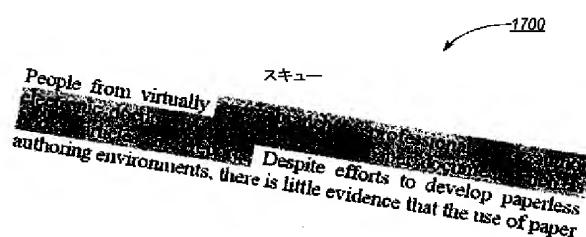
【図14】



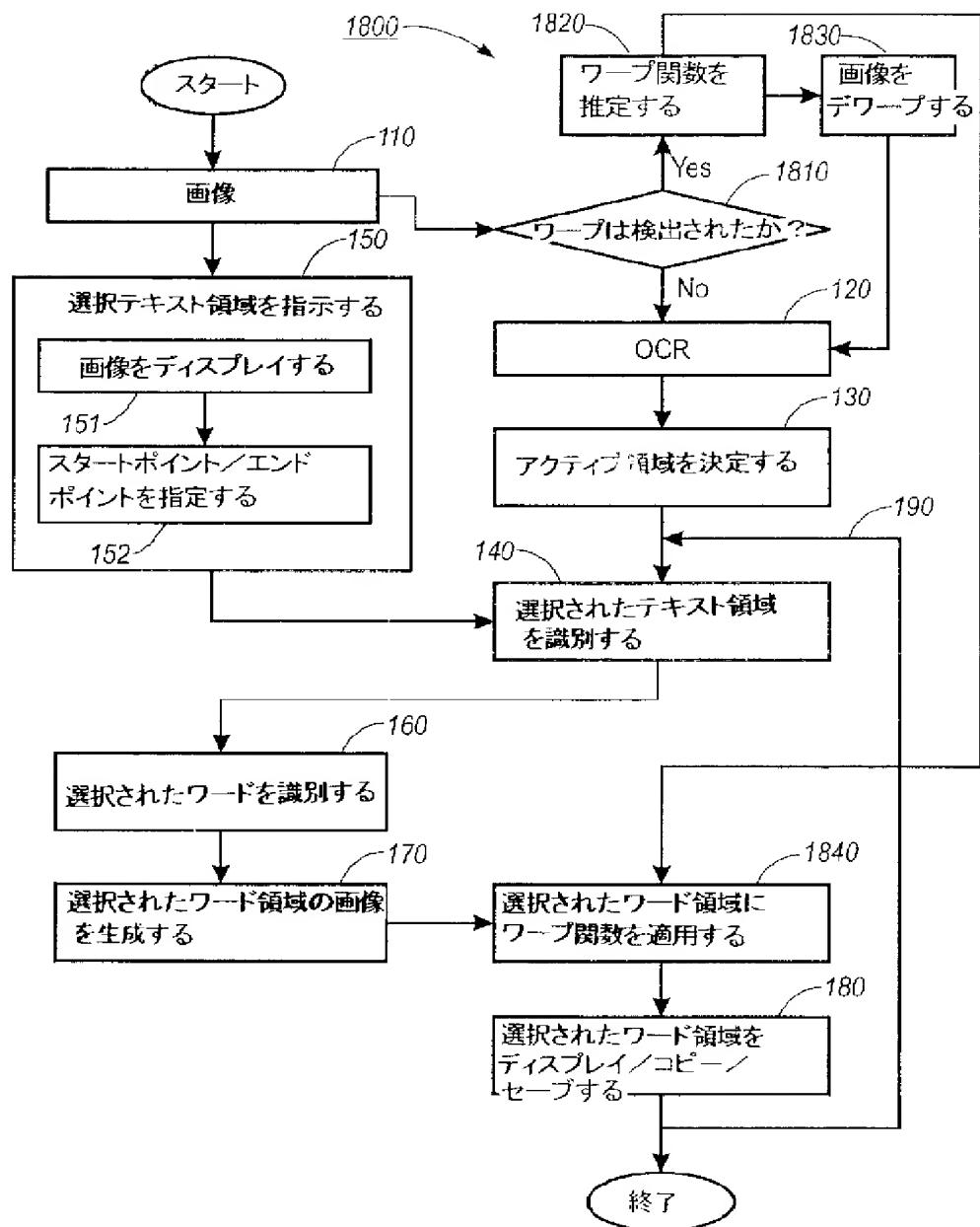
【図15】



[図17]



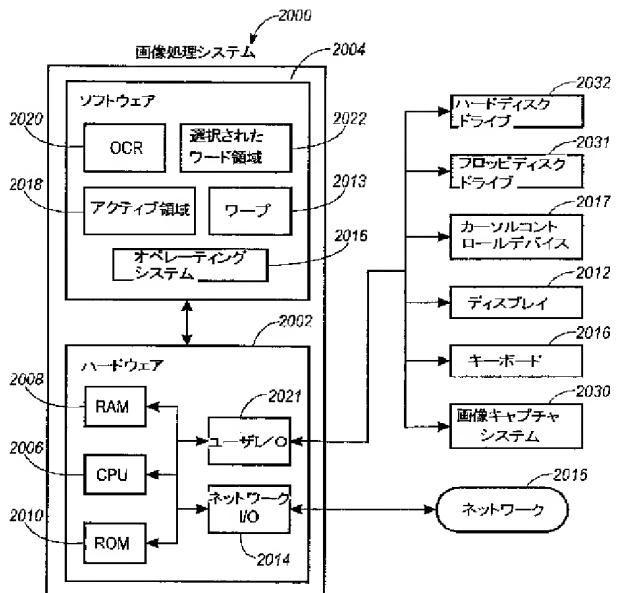
【図18】



【図19】

every branch of professional work author  
electronic documents while referring to paper documents such as  
books, articles and reports.

【図20】



## フロントページの続き

(72)発明者 クリストファー アール. ダンス  
イギリス国 シービー2 2エイチユー  
ケンブリッジ トランピントン スコット  
ダウン ロード 24

(72)発明者 スチュアート エー. ティラー  
イギリス国 シービー1 3エスエス ケ  
ンブリッジ リッチフィールド ロード